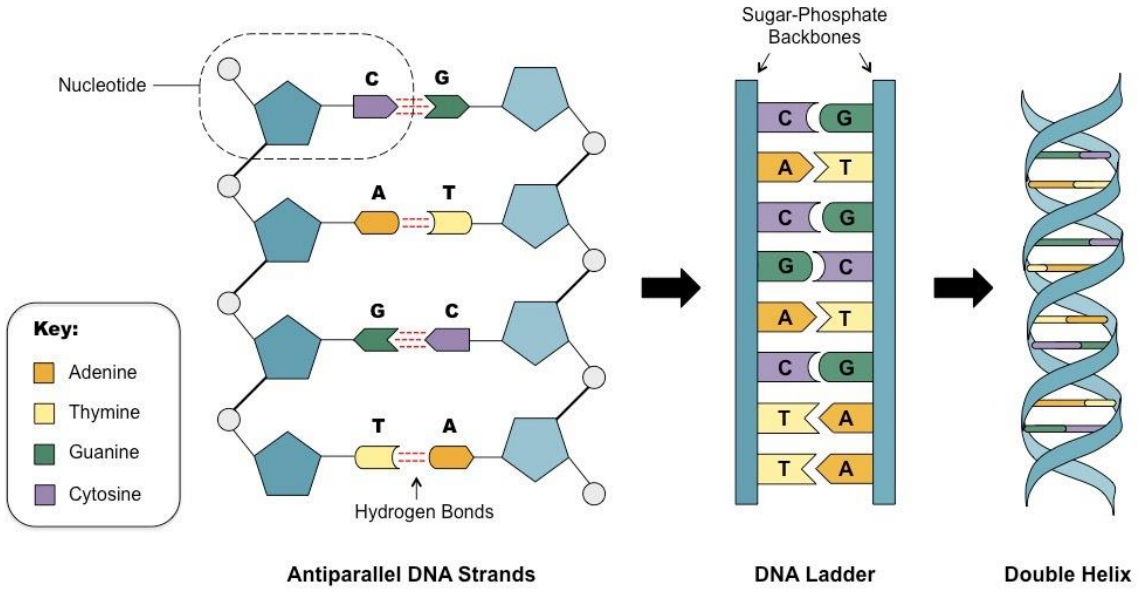
Tetramer-Dependence of DNA Conformation

ZOE WEFERS

MENTOR: DR. WILMA OLSON

DNA: A Biological View

- DNA is a biological molecule that encodes genetic information
- 4 types of nucleic acid bases, adenosine (A), cytosine (C), guanosine (G), thymidine (T)
- Nucleic acid bases form base pairs
  - A pairs with T and C pairs with G
- Read one strand from 5' to 3' - ex: CACGACTT
- Double helix structure - stiff
  - Bending and twisting affects elastic energy

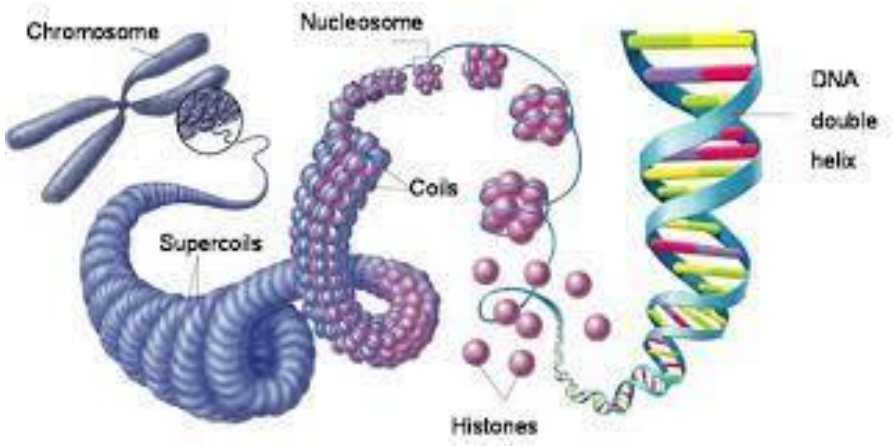


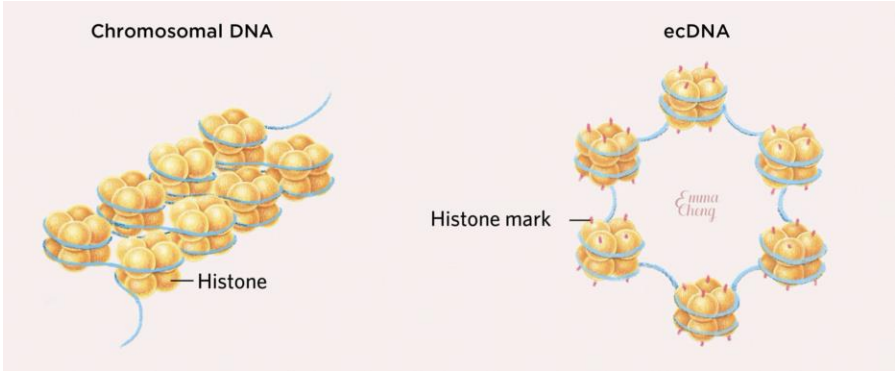
https://ib.bioninja.com.au/standard-level/topic-2-molecular-biology/26-structure-of-dna-and-rna/dna-structure.html

DNA: The Important of Folding

- Packaged in nucleus by wrapping around histone proteins
- DNA conformation can influence gene regulation
  - Ex: transcription
- Extrachromosomal circular DNA is found in some cancers
  - Why we study DNA minicircles

[4]



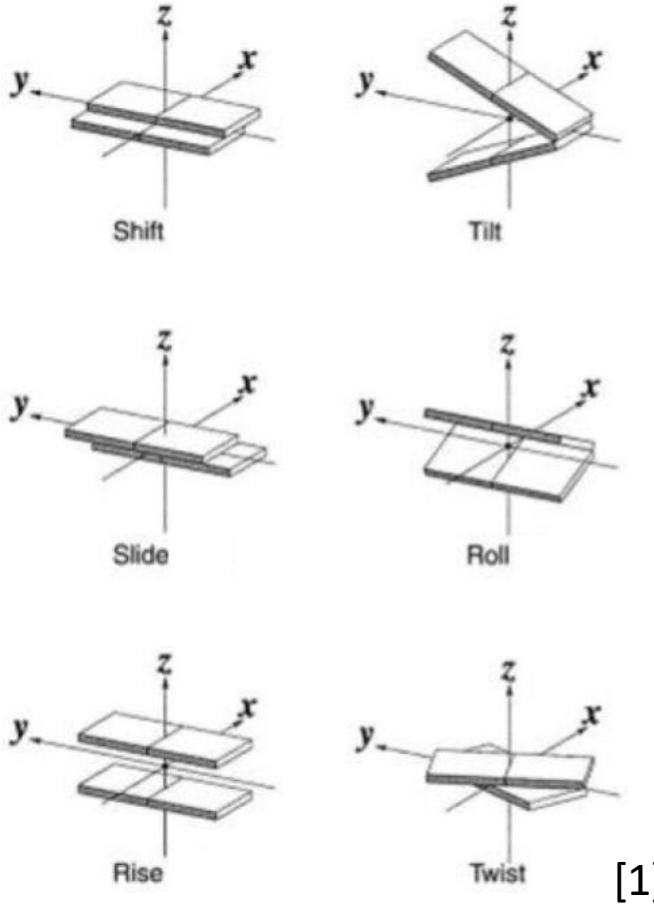


https://www.online-sciences.com/biology/packaging-of-dna-genome-chromosomal-proteins-dna-in-prokaryote
eukaryotes/attachment/packaging-of-dna-88/%E2%80%8B
https://www.the-scientist.com/features/cancer-may-be-driven-by-dna-outside-of-chromosomes-68590

DNA: A Mathematical View

- Model each base pair as rigid rectangle
- Orientation between adjacent base pairs described represented by 6 "base pair step parameters": Tilt, Roll, Twist, Shift, Slide, Rise
  - $\underline{\mathbf{p}} = (\theta_1, \theta_2, \theta_3, \rho_1, \rho_2, \rho_3)$
  - Intrinsic parameter, denoited $\underline{\mathbf{p}}_0$ = Rest state step parameters
- Stiffness matrix ($\mathbf{F}$) = covariance matrix inverse

$$\mathbf{F}^{-1} = \begin{bmatrix} \langle\theta_1^2\rangle - \langle\theta_1\rangle^2 & \langle\theta_1\theta_2\rangle - \langle\theta_1\rangle\langle\theta_2\rangle & \dots & \dots & \dots & \langle\theta_1\rho_3\rangle - \langle\theta_1\rangle\langle\rho_3\rangle \\ \langle\theta_1\theta_2\rangle - \langle\theta_1\rangle\langle\theta_2\rangle & \langle\theta_2^2\rangle - \langle\theta_2\rangle & \dots & \dots & \dots & \langle\theta_2\rho_3\rangle - \langle\theta_2\rangle\langle\rho_3\rangle \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \langle\theta_1\rho_3\rangle - \langle\theta_1\rangle\langle\rho_3\rangle & \langle\theta_2\rho_3\rangle - \langle\theta_2\rangle\langle\rho_3\rangle & \dots & \dots & \dots & \langle\rho_3^2\rangle - \langle\rho_3\rangle \end{bmatrix}$$



[1]

Elastic Energy

$$\underline{p}^i = (\theta_1^i, \theta_2^i, \theta_3^i, \rho_1^i, \rho_2^i, \rho_3^i)$$

Step Parameters for the ith step

$$E^i = \frac{1}{2}(\underline{p}^i - \underline{p}_0^{\,i})^T \mathbf{F}^i (\underline{p}^i - \underline{p}_0^{\,i})$$

Elastic Energy for the ith step

$$E = \sum_{i=1}^{N-1} E^i$$

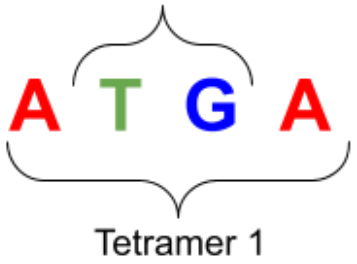
Total Elastic

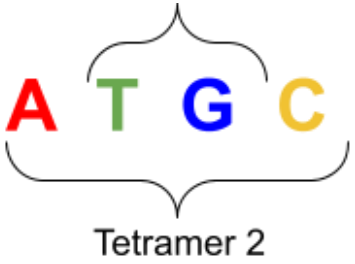
$$\frac{\partial E^i}{\partial \underline{p}^i} = \mathbf{F}_s^i (\underline{p}^i - \underline{p}_0^{\,i})$$

Derivative of Elastic Energy

What is a Sequence Dependent Model?

- Find intrinsic parameter and stiffness matrices using either high resolution structural data or molecular dynamics simulation data
- Dimer-dependent Model: Intrinsic parameters depends on the base pairs in the step
  - ie intrinsic parameters for AA step different than GC step
- Tetramer Dependent model: Intrinsic parameters depends on tetramer centered at the step
  - New evidence suggests intrinsic parameter are affected also by flanking base pairs

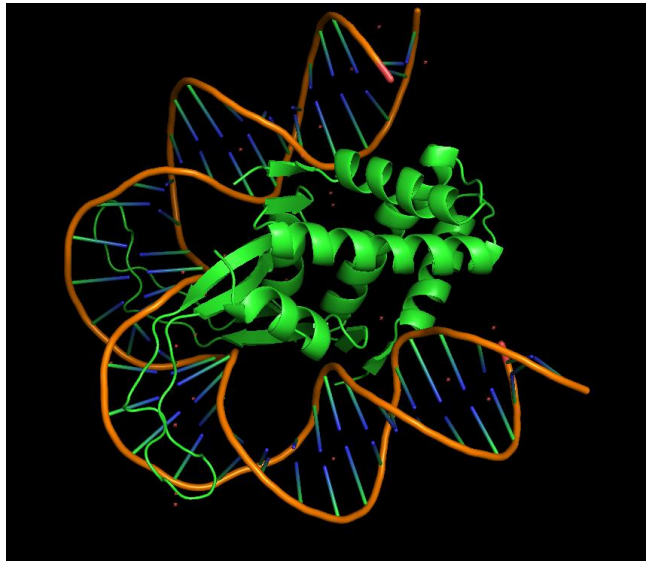


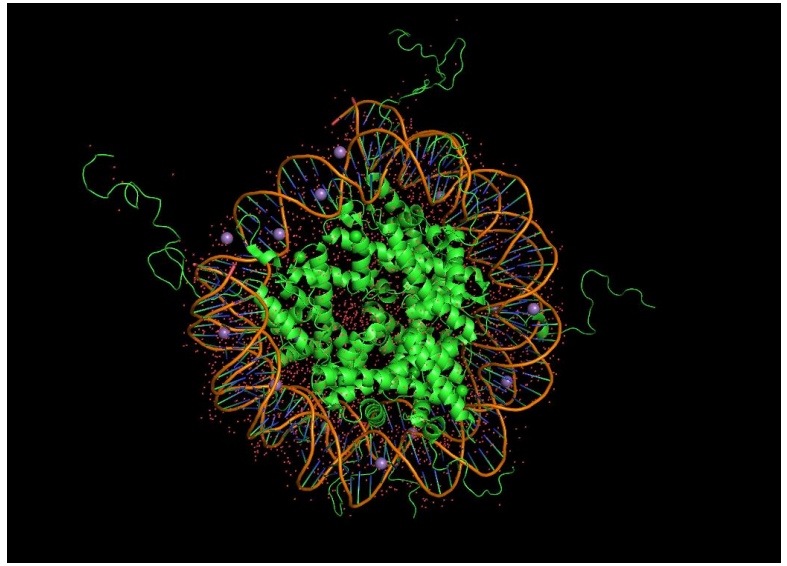


- Models we used: IdealDNA (sequence-independent), Olson1998 (dimer), Cohen2017_dim (dimer), Cohen2017_tet (tetramer)

emDNA Software – Current Functionalities

- Core Idea: DNA will move to achieve least elastic energy
  - Can be modeled by gradient descent-like algorithm
- **User has the option to use dimer-dependent models or sequence independent model**
  - **But no tetrameric functionality**





My Goals:

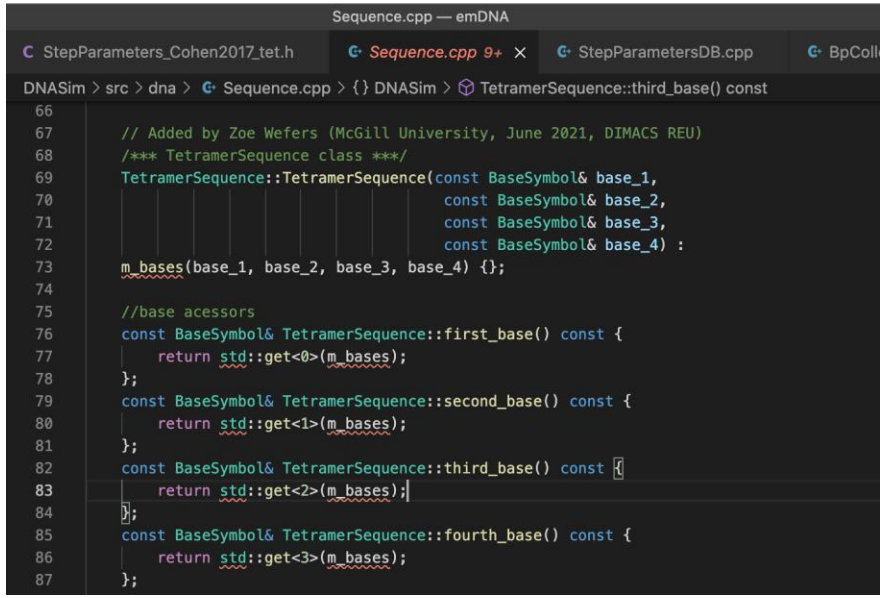
1) Adapt emDNA to allow user to choose to tetramer-dependent, dimer-dependent or sequence independent model to optimize elastic energy

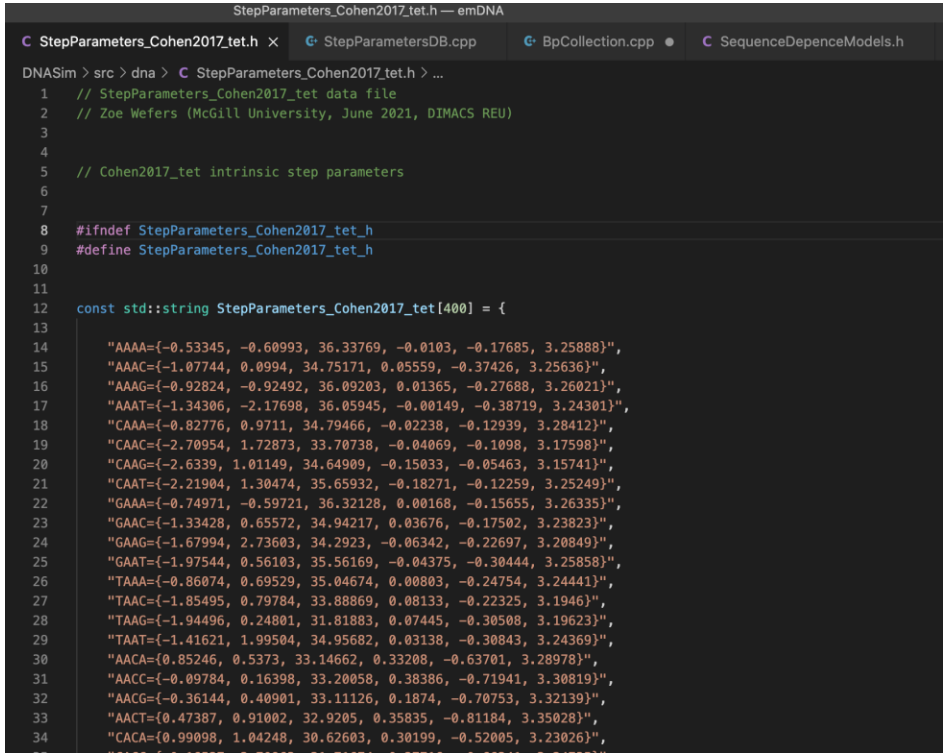
2) Compare the intrinsic parameters of new Cohen2017 sequence dependent models

3) Use the modified program to see how using different sequence dependent models changes the minimum energy conformations achieved

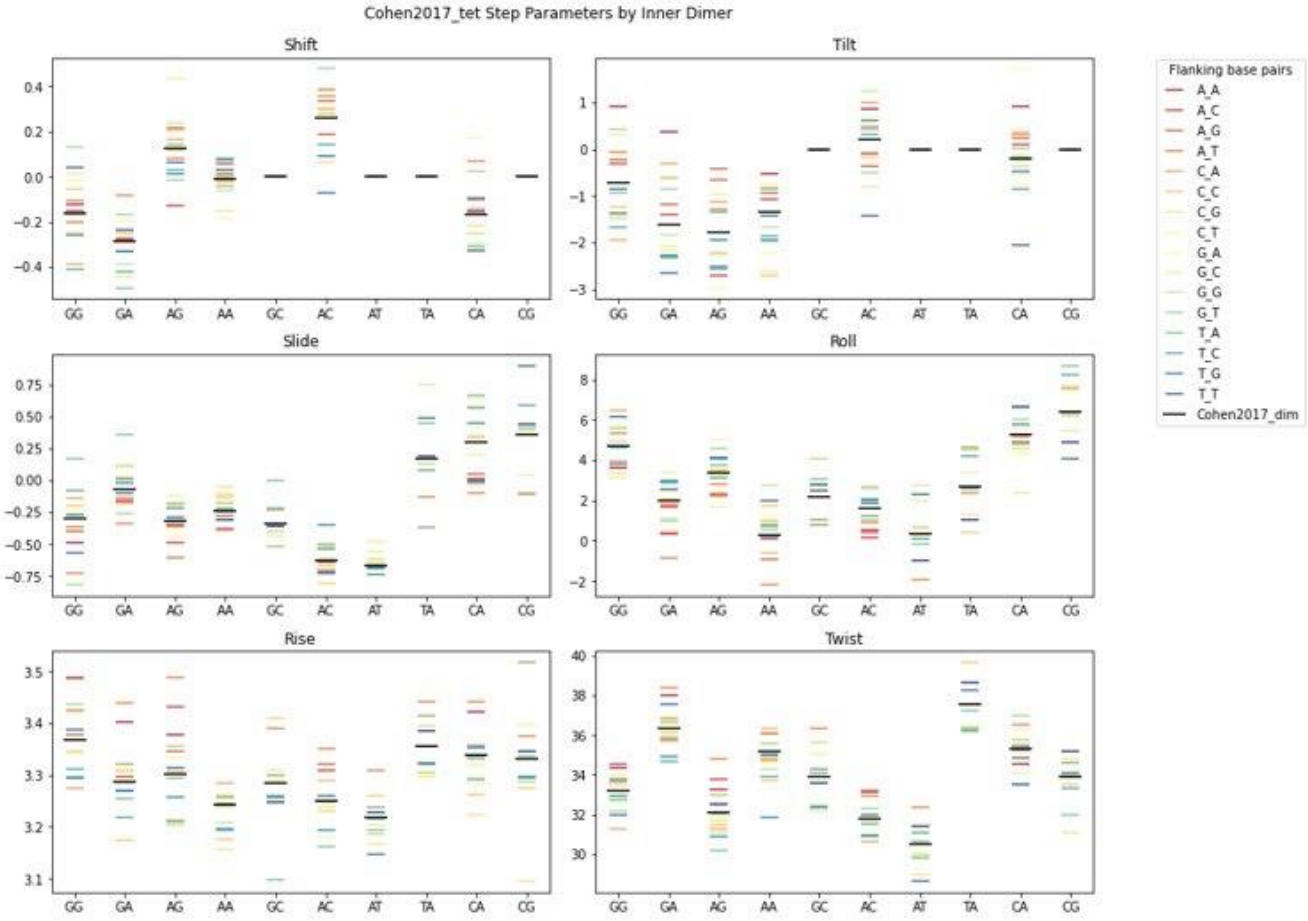
Goal 1: Adapting emDNA

- Successfully implemented the tetrameric functionality!



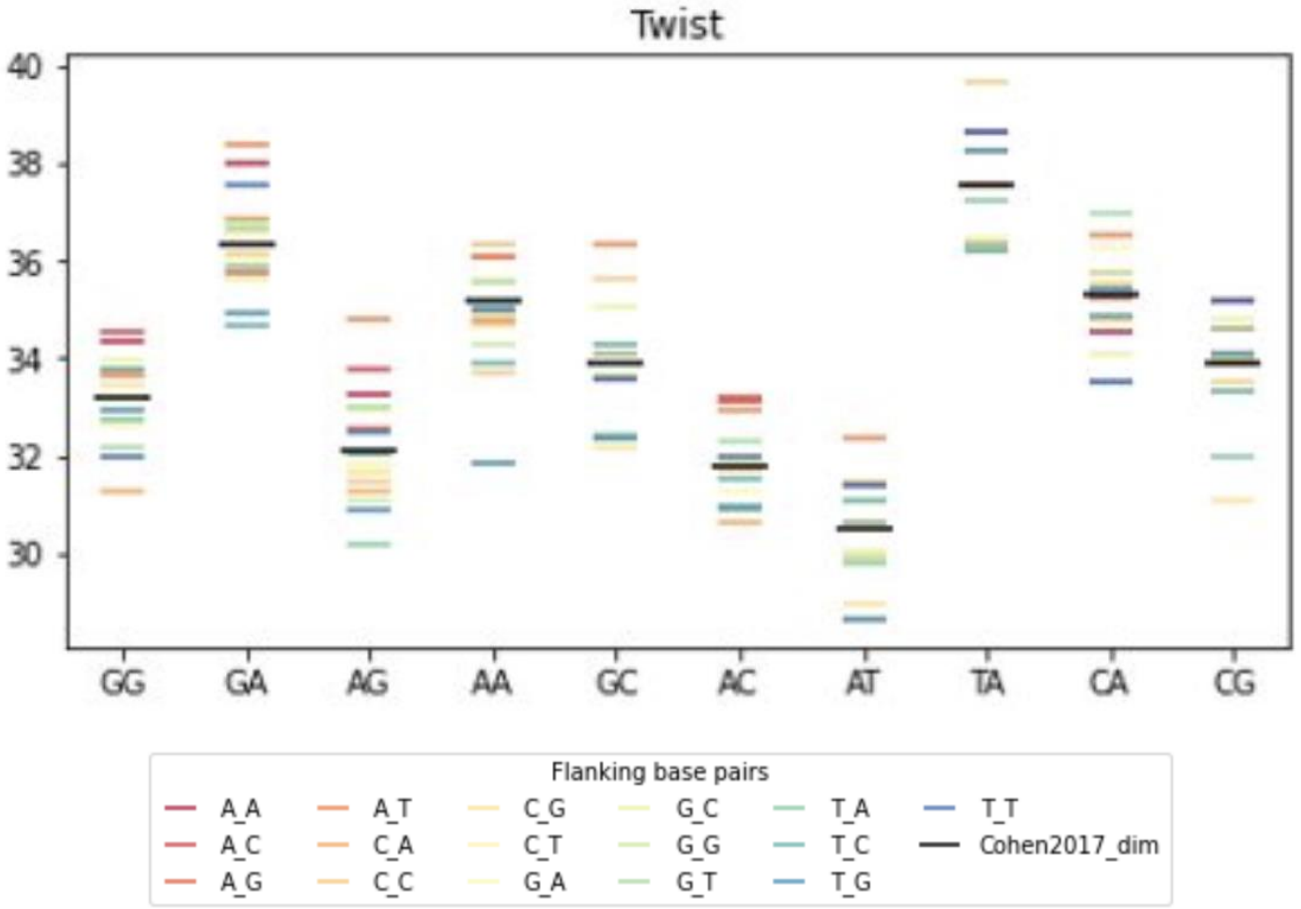


Goal 2: Comparing Cohen2017 models



- Significant effect of flanking base pairs provides evidence that potential differences in optimized configuration of are in fact caused by dimer vs tetramer models
  - Not by other computational factors

Goal 2: Comparing Cohen2017 models



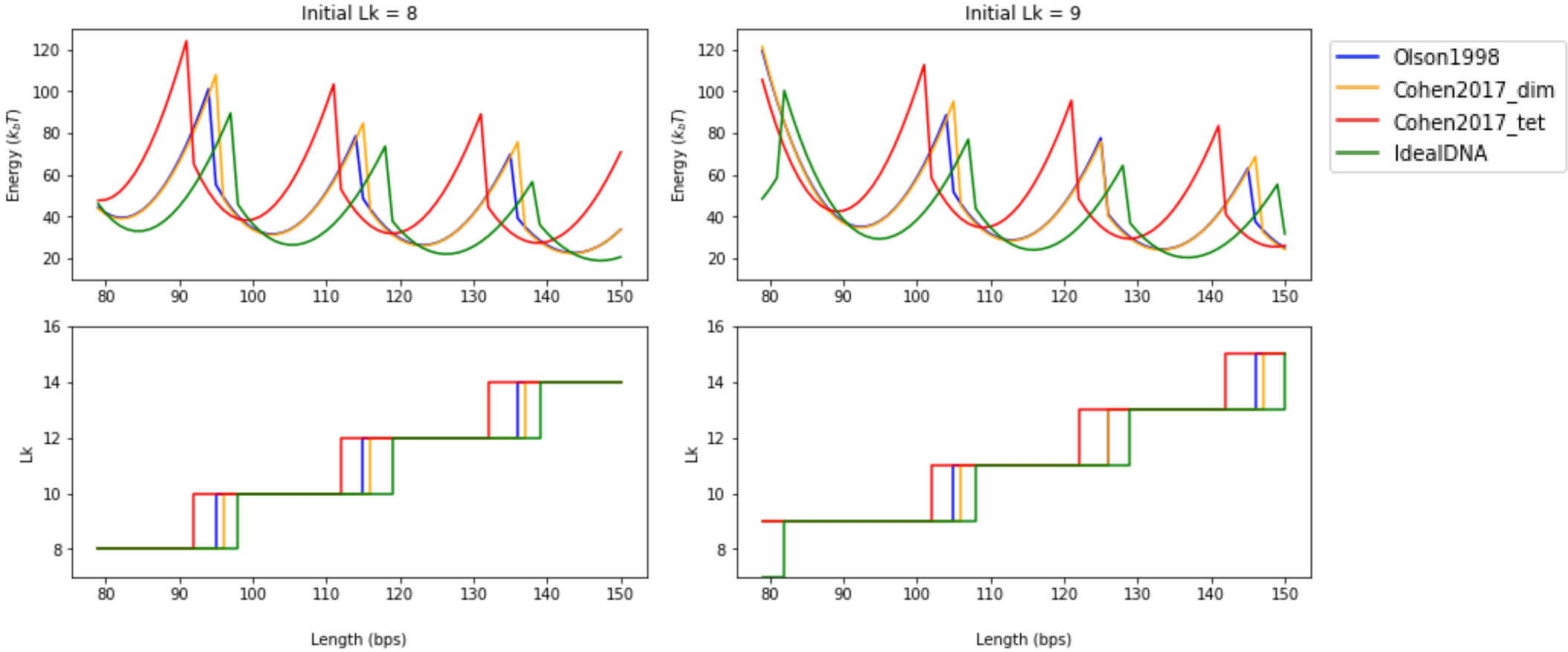
- Significant effect of flanking base pairs provides evidence that potential differences in optimized configuration of are in fact caused by dimer vs tetramer models
  - Not by other computational factors

Goal 3: Optimizing Minicircles

Experimental Setup:

- Choose tetramer that is 1) easily repeatable and 2) intrinsic parameters vary significantly from intrinsic parameters of inner dimer
  - AAAA - 1.5^{o} increase in intrinsic twist for AAAA tetramer compared to AA dimer in Cohen2017 models
- Optimize DNA Minicircles of lengths 80-150 base pairs and with linking number 8 and 9

Final Energy and Final Linking Number



Conclusions

Goal 1: Created new tool that expands the possibilities for exploring sequence dependence in DNA structure

Goal 2: Provided further evidence that base pairs which flank inner dimer impact intrinsic parameters

Goal 3: Results of optimization confirm that the scope of sequence dependent models has non-trivial effects on final conformation of DNA minicircles

Next Steps

- Use higher resoluter data to recompute Cohen2017 models
- Optimize DNA minicircles with more varied sequences
  - Maybe use repeating AG, AAGG, AAAGGG, etc.
- First and last base pair steps in in a linear/open piece DNA sequence do not correspond to a specific tetramer
  - Process trimer data and include it in tetramer-dependent models

References

1. Lu, 2003, '3DNA: A Software package for the Analysis, Rebuilding and Visualization of Three-Dimensional Nucleic Acid Structures', *Nucleic Acid Research*

2. Olson, W. K. (1998). DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proceedings of the National Academy of Sciences*. Published. https://doi.org/10.1073/pnas.95.19.11163

3. Clauvelin, Olson, 2021, 'Synergy between Protein Positioning and DNA Elasticity: Energy Minimization of Protein Decorated Minicircles', *Journal of Physical Chemistry*

4. Irobalieva, R. N., Fogg, J. M., Catanese, D. J., Sutthibutpong, T., Chen, M., Barker, A. K., Ludtke, S. J., Harris, S. A., Schmid, M. F., Chiu, W., & Zechiedrich, L. (2015). Structural diversity of supercoiled DNA. *Nature Communications*, *6*(1). https://doi.org/10.1038/ncomms9440

5. Pasi, M., Maddocks, J. H., Beveridge, D., Bishop, T. C., Case, D. A., Cheatham, T., Dans, P. D., Jayaram, B., Lankas, F., Laughton, C., Mitchell, J., Osman, R., Orozco, M., Pérez, A., Petkevičiūtė, D., Spackova, N., Sponer, J., Zakrzewska, K., & Lavery, R. (2014). μABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Research*, *42*(19), 12272–12283. https://doi.org/10.1093/nar/gku855

Acknowledgments

Thank you to the DIMACS REU 2021 program at Rutgers University, supported by NSF grant CCF-1852215. And thank you to Dr. Olson and Robert Young for their guidance!